

Ekstraksi dan Seleksi Fitur untuk Klasifikasi Sel Epitel dengan Sel Radang pada Citra Pap Smear

Rahadian Kurniawan

Jurusan Teknik Informatika,
Universitas Islam Indonesia
Yogyakarta, Indonesia
rahadiankurniawan@uii.ac.id

Dinda Eling Kartikaning Sasmito

Magister Teknik Informatika,
Universitas Islam Indonesia
Yogyakarta, Indonesia
dindaelingsasmito@gmail.com

Fajar Suryani

Magister Teknik Informatika,
Universitas Islam Indonesia
Yogyakarta, Indonesia
fajarsuryani@gmail.com

Abstrak— Penelitian ini dilakukan seleksi fitur menggunakan *Fisher Criterion*, sedangkan pada proses klasifikasi data menggunakan algoritma *Backpropagation* terhadap 16 fitur yang terlebih dahulu diekstrak dari citra Pap smear. Adapun ke-16 fitur yang digunakan dibagi menjadi 3 kategori, yaitu: Fitur bentuk, Fitur tekstur, dan Fitur intensitas warna. Pada naskah ini terdapat 2 tahap utama, yaitu: 1) Ekstraksi Fitur; dan 2) Seleksi Fitur. Penelitian ini bertujuan menganalisis kinerja seleksi fitur pada klasifikasi data dan mencari fitur yang secara signifikan mempengaruhi klasifikasi sel epitel dengan sel radang. Sebagai pembandingan, penelitian ini juga membandingkan hasil seleksi fitur antara *Fisher Criterion* dengan *Feature Subset Selection*. Hasil yang diperoleh dari proses perbandingan tersebut menunjukkan kesamaan fitur yang secara signifikan mempengaruhi proses klasifikasi sel radang dengan sel epitel. Tingkat akurasi klasifikasi pada penelitian ini adalah 92.5%.

Keywords: *Fisher Criterion*, *Backpropagation*, *Kanker Serviks*

I. PENDAHULUAN

Tes Pap smear adalah prosedur *screening* manual yang digunakan untuk mendeteksi kanker serviks atau perubahan prakanker dengan melakukan penilaian berdasarkan warna, bentuk, dan tekstur pada sel epitel atau daerah *nuclei* (inti sel). Pap Smear merupakan suatu metode pemeriksaan sel-sel yang diambil dari leher rahim dan kemudian diperiksa di bawah mikroskop. Interpretasi visual citra Pap smear manual memiliki banyak keterbatasan, diantaranya masalah prosedur pewarnaan, terdapat darah yang berlebihan, lendir, bakteri dan adanya sel radang yang menyebabkan proses pengenalan sel menjadi sulit, lama dan rawan kesalahan prosedur. Oleh karena itulah perlu adanya analisis secara otomatis berbantuan komputer untuk membantu memudahkan proses interpretasi tersebut.

Dalam beberapa tahun terakhir berbagai upaya telah dilakukan oleh beberapa peneliti untuk berkontribusi pada analisis dan interpretasi otomatis berbantuan komputer dengan citra Pap Smear. Beberapa penelitian yang dilakukan untuk menginterpretasi citra Pap smear antara lain adalah: melakukan deteksi letak *nuclei* dengan metode *Morphological Reconstruction* dan *Clustering* [1], *Fuzzy C-Means Clustering* [2], dan *Edge detectors* [3], [4]. Selain itu, penelitian sejenis yang dilakukan untuk melakukan segmentasi *nuclei* antara lain: *Deformable Templates* [5], *Pixel Classification Schemes* [6] *Morphological Operation and Watershed Transformation* [7]. Beberapa penelitian tersebut menunjukkan hasil yang

menjanjikan dalam tujuannya masing-masing. Akan tetapi, belum banyak penelitian yang dilakukan untuk melakukan seleksi fitur untuk mengklasifikasikan antara sel epitel dengan sel radang pada citra Pap smear.

Pada banyak kasus, dokter mengalami kesulitan untuk membedakan sel epitel dengan sel radang dalam pembacaan manual melalui mikroskop. Pada beberapa penelitian sebelumnya [1],[2], masih terdapat kesalahan interpretasi karena komputer masih mengenali sel radang sebagai sel epitel. Banyaknya kesalahan interpretasi ini karena adanya kemiripan bentuk, tekstur dan intensitas warna antara sel epitel dan sel radang. Gambar 1 menunjukkan kemiripan antara sel epitel dan sel radang. Pada Gambar. 1(a) terlihat bahwa sel radang (B) memiliki bentuk dan ukuran yang hampir sama dengan sel epitel (A). Gambar 1 (b) menunjukkan peta warna dari Gambar 1(a). Pada Gambar 1 (b) warna biru menunjukkan intensitas rendah, sedangkan warna merah menunjukkan intensitas tinggi. Dari gambar tersebut terlihat bahwa sel epitel (C) memiliki intensitas warna yang sama bahkan cenderung lebih tinggi (terang) daripada sel radang (D).

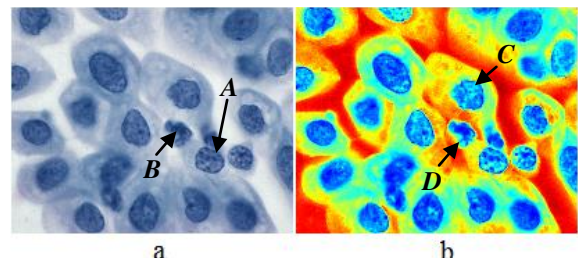


Fig. 1. (a) Bagian dari citra Papsmear, (b) Peta warna dari Gambar 1.(a).

Untuk membedakan sel epitel dan sel radang dapat dilakukan dengan menggunakan algoritma klasifikasi. Pada umumnya algoritma klasifikasi menggunakan semua fitur yang terdapat pada data untuk membangun sebuah model, padahal tidak semua fitur tersebut relevan terhadap hasil klasifikasi dan tentu mempengaruhi waktu pemrosesan, apalagi jika data yang digunakan berdimensi besar. Salah satu solusi yang digunakan untuk mengatasi masalah tersebut adalah dengan menggunakan seleksi fitur. Penelitian sebelumnya[8] menunjukkan bahwa seleksi fitur terbukti mampu memilih fitur yang relevan dengan performa waktu pemrosesan data yang baik.

Penelitian yang akan dilakukan merupakan salah satu sub-bagian dari penelitian lain yaitu untuk mendeteksi kanker serviks pada citra Pap smear. Penelitian pada naskah ini hanya berfokus untuk melakukan proses ekstraksi terhadap fitur bentuk, tekstur, dan intensitas warna pada setiap sel epitel dan sel radang dalam citra Pap smear serta melakukan seleksi fitur yang relevan untuk klasifikasi sel epitel dan sel radang. Proses ekstraksi fitur dilakukan terhadap 16 fitur antara lain: *Circularity, Entropy, Uniformity, Mean, Standard deviation, Eccentricity, Semi Major and Minor Axis Length, Equivalent Diameter, Skewness, Smoothness, Average Radius, Compactness, Foreground-Background contrast in red, Foreground-Background contrast in green, and Foreground-Background contrast in blue*. Dari ke-16 fitur tersebut kemudian akan dicari fitur yang paling berpengaruh untuk klasifikasi sel epitel dan sel radang. Penelitian ini mengusulkan metode *Fisher Criterion* dilanjutkan dengan *Backpropagation* sebagai metode yang digunakan untuk proses seleksi fitur. Output penelitian ini adalah berupa rekomendasi fitur terbaik untuk klasifikasi sel epitel dan sel radang yang dapat digunakan untuk penelitian selanjutnya. Pada naskah ini terdapat 2 tahap utama, antara lain: 1) Ekstraksi Fitur; dan 2) Seleksi Fitur. Penjelasan lebih lanjut akan dijelaskan pada subbagian selanjutnya.

II. METODE

A. Ekstraksi Fitur

Tahap ekstraksi fitur merupakan tahap setelah proses segmentasi citra pada penelitian sebelumnya [7]. Fitur yang akan di ekstrak pada tahap ini terdiri dari 16 fitur yang diambil dari beberapa penelitian sebelumnya [9], [10], [11]. Fitur tersebut dikelompokkan menjadi 3 berdasarkan warna, bentuk, dan tekstur. Berikut penjelasan detail mengenai 3 kelompok fitur tersebut.

1. Fitur Bentuk

Sel epitel dan sel radang memiliki bentuk yang hampir mirip. Akan tetapi, sel epitel memiliki bentuk yang lebih besar dan lebih teratur dari pada sel radang. Gambar 2 mengilustrasikan perbedaan bentuk antara sel epitel dan sel radang. Pada Gambar 2(c) dan (d) terlihat hasil segmentasi citra menunjukkan perbedaan antara sel epitel dan sel radang dalam hal ukuran (diameter), jari-jari, maupun keteraturan bentuk.

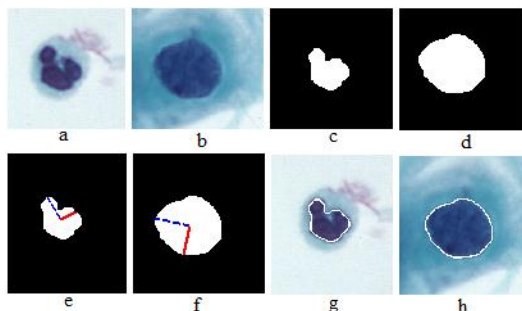


Fig. 2. (a) Sel radang, (b) Sel Epitel, (c) Segmentasi sel radang, (d) Segmentasi sel epitel, (e) Segmentasi sel radang dengan penanda *Major* dan *Minor Axis*, (f) Segmentasi sel epitel dengan penanda *Major* dan *Minor Axis*, (g) Sel radang dengan *boundary marker*, dan (h) Sel epitel dengan *boundary marker*.

Fitur bentuk yang akan di ekstrak pada tahap ini adalah: *Circularity, Eccentricity, Semi Major and Minor Axis Length, Equivalent Diameter, Average Radius, Compactness*.

Pada Gambar 2(e) terlihat bahwa sumbu *major* (garis putus-putus) cenderung jauh lebih panjang daripada sumbu *minor* (garis sambung). Sedangkan pada Gambar 2(f) terlihat bahwa sumbu *major* (garis putus-putus) cenderung sama panjang atau hanya sedikit perbedaan daripada sumbu *minor* (garis sambung).

Untuk menghitung semi-*Major axis* yang dapat dilakukan adalah mencari koordinat terjauh antara perimeter hasil segmentasi dengan *centroid*. Selanjutnya, untuk dapat menemukan semi-*Minor axis* adalah dengan mencari koordinat titik tegak lurus dengan semi-*Major axis*. Untuk mendapatkan titik yang tegak lurus tersebut dapat diperoleh dengan,

$$\begin{aligned} dx &= x_2 - x_1; \\ dy &= y_2 - y_1; \\ (-d_y, d_x) \text{ and } (d_y, -d_x) \end{aligned} \quad (1)$$

dimana x_1 dan y_1 adalah koordinat piksel *centroid*, dan x_2 dan y_2 adalah koordinat titik semi-*Major axis*. Akan tetapi, kedua koordinat tegak lurus tersebut masih sama panjang dengan semi-*Major axis* (Gambar 3 (d)) sehingga koordinat masih berada di luar perimeter. Untuk mendapatkan koordinat perpotongan antara sumbu minor dan koordinat dari perimeter dapat menggunakan persamaan,

$$\begin{aligned} m &= (y_2 - y_1)/(x_2 - x_1); \\ b &= -((m \times dx) - dy); \\ y &= m + b \end{aligned} \quad (2)$$

dimana x_1 dan y_1 adalah koordinat *centroid*, dan x_2 dan y_2 adalah koordinat titik semi-*major axis*. Selanjutnya, dari 2 semi-*minor* yang didapat kemudian dibandingkan panjangnya (Gambar. 3 (d) garis hijau (b_1) dan garis merah muda (b_2)). Koordinat dengan jarak terpendek dari *centroid* akan dianggap sebagai semi-*minor axis* (Gambar. 3 (f)).

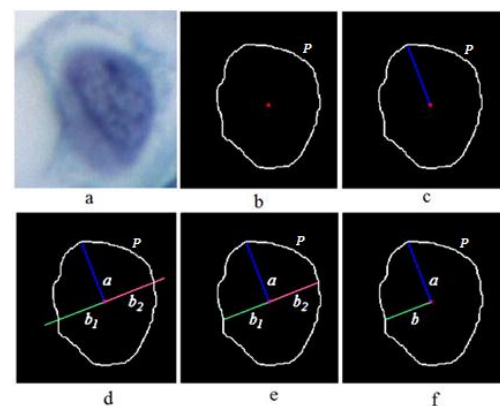


Fig. 3. (a) *nuclei*, (b) *Nuclei boundary*, (c) *Nuclei boundary* dengan semi-*major axis*, (d) Sumbu tegak lurus 90 derajat dari semi-*major axis*, (e) *nuclei boundary* dengan *minor axis*, (f) *nuclei boundary* dengan semi-*minor axis* (b).

Adapun untuk mendapatkan ekstraksi fitur yang lain (*Circularity, Eccentricity, Equivalent Diameter, Average Radius, Compactness*.) Tabel 1 menunjukkan persamaan-persamaan matematika yang digunakan.

TABLE I. FITUR BENTUK

| Fitur Bentuk | |
|-------------------------------|---|
| <i>Semi minor-Axis length</i> | b = jarak terpendek <i>centroid</i> dengan <i>boundary nuclei</i> |
| <i>Semi major-Axis length</i> | a = jarak terpanjang <i>centroid</i> dengan <i>boundary nuclei</i> |
| <i>Average Radius</i> | Avr = rata-rata jarak <i>centroid</i> dengan <i>boundary nuclei</i> |
| <i>Eccentricity</i> | $c = \sqrt{a^2 - b^2}$; $e = c/a$ |
| <i>Equivalent Diameter</i> | $ED = \frac{4 \times Area}{\pi}$ |
| <i>Perimeter</i> | P = Jumlah pixel pada <i>boundary nuclei</i> |
| <i>Circularity</i> | $C = \frac{4\pi \times Area}{P^2}$ |
| <i>Compactness</i> | $Cp = P^2 / Area$ |

TABLE II. FITUR TEKSTUR

| Fitur Tekstur | |
|--|---|
| <i>Mean (First Momment)</i> | $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ dimana n adalah banyaknya pixel dalam citra, dan x_i adalah nilai intensitas warna pada tiap pixel |
| <i>Stdev (Second Moment)</i> | $S_{N-1} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ |
| <i>Skewness (Third Moment)</i> | $skew = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)S_{N-1}^3}$ |
| <i>Smoothness</i> | $R = 1 - \frac{1}{1 + S_{N-1}^2}$ |
| <i>Entropy</i> | $H(x) = - \sum_{i=1}^n p(z_i) \times \log_2 p(z_i)$ |
| <i>Uniformity</i> | $U = \sum_{i=0}^n p^2(z_i)$ |
| z_i adalah nilai intensitas pada setiap pixel, dan $p(z)$ adalah jumlah pixel untuk setiap intensitas pada histogram dibagi dengan jumlah pixel. | |

2. Fitur Tekstur

Analisis fitur tekstur dari hasil segmentasi dilakukan berdasarkan analisis statistik *histogram* intensitas pada 3 layer warna (*red, green, blue*). Sel epitel dan sel radang memiliki tekstur yang cenderung berbeda. Pada Gambar 4, terlihat bahwa sel radang memiliki persebaran intensitas warna yang agak lebar (Gambar 4(b)) daripada sel epitel (Gambar 4(d)). Hal tersebut menunjukkan intensitas warna pada sel epitel lebih homogen daripada sel radang yang dalam hal ini berarti sel epitel memiliki tekstur yang lebih halus daripada tekstur sel radang.

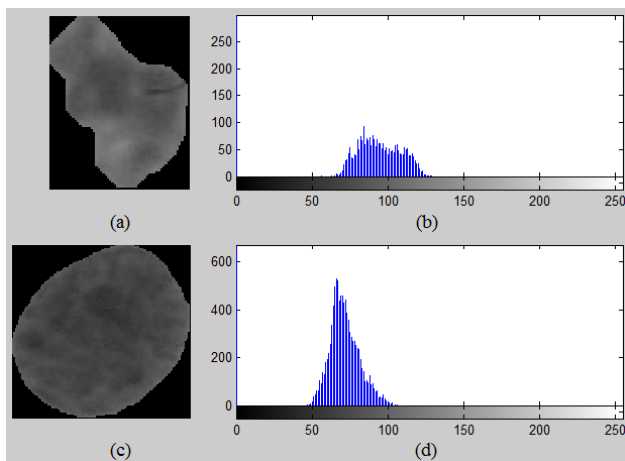


Fig. 4. (a) Sel radang (*red layer*), (b) *histogram* sel radang (a), (c) Sel epitel (*red layer*), (d) *histogram* sel epitel (c).

Adapun untuk mendapatkan ekstraksi fitur tekstur (*Entropy, Uniformity, Mean, Standard deviation, Skewness, Smoothness*) berikut Tabel 2 menunjukkan persamaan-persamaan matematika yang digunakan.

3. Fitur Intensitas Warna

Perbedaan yang cukup signifikan antara sel epitel dan sel radang adalah pada sitoplasma. Sitoplasma pada sel epitel cenderung besar sehingga kontras dengan latar belakang citra tinggi. Sedangkan pada sel radang sitoplasma biasanya kecil bahkan tidak terlihat, sehingga kontras dengan latar belakang citra rendah. Gambar 5 menunjukkan ilustrasi perbedaan besar sitoplasma sel epitel (B) dan sel radang (A).

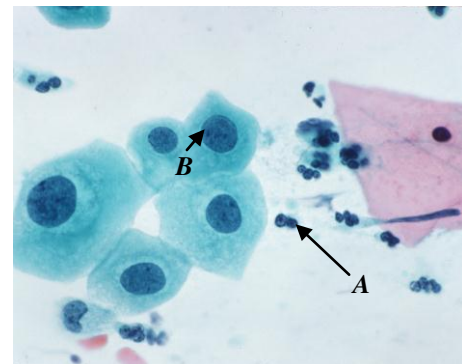
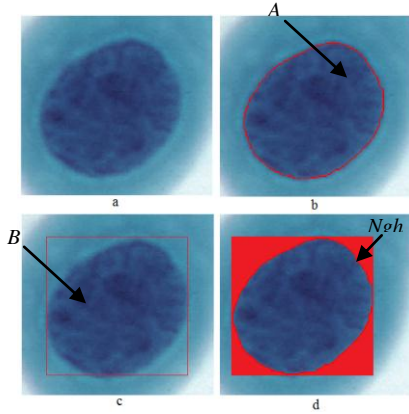


Fig. 5. Ilustrasi perbedaan sitoplasma sel epitel dan sel radang.

Untuk mendapatkan nilai kontras sitoplasma, berikut tabel 3 menunjukkan persamaan-persamaan matematika yang digunakan, sedangkan Gambar 6 menunjukkan ilustrasi ekstraksi fitur intensitas warna. Setelah mendapatkan semua nilai pada fitur bentuk, tekstur, dan intensitas warna, langkah berikutnya adalah melakukan seleksi fitur. Adapun proses ekstraksi fitur dilakukan dengan menggunakan software Matlab.

TABLE III. FITUR INTENSITAS

| Fitur Intensitas | |
|---|------------------------------------|
| Foreground-Background contrast in red | $nghR = m_{red}^B - m_{red}^A$ |
| Foreground-Background contrast in green | $nghG = m_{green}^B - m_{green}^A$ |
| Foreground-Background contrast in blue | $nghB = m_{blue}^B - m_{blue}^A$ |

Fig. 6. (a) Sel epitel, (b) Sel epitel dengan nuclei boundary marker, (c) Boundingbox dari hasil segmentasi, (d) Wilayah ngh ($A^c \cap B$).

B. Seleksi Fitur

1. Fisher Criterion

Pada tahap ini akan dipilih fitur-fitur yang memiliki tingkat signifikan yang tinggi untuk mengklasifikasikan sel epitel maupun sel radang terhadap 16 fitur yang digunakan. Seleksi fitur berdasarkan cara mengevaluasi kualitas fiturnya dapat dibedakan menjadi 2 kategori. Kategori pertama yaitu seleksi fitur dengan pendekatan *filter*. Pada pendekatan ini, evaluasi fitur terbaik adalah berdasarkan karakteristik data itu sendiri dan tidak bergantung pada algoritma klasifikasi. Kategori yang kedua yaitu seleksi fitur dengan pendekatan *wrapper*. Pada pendekatan wrapper, evaluasi fitur memanfaatkan sebuah algoritma mining sehingga proses komputasinya lebih lama dibandingkan dengan seleksi fitur dengan pendekatan *filter*. Salah satu metode yang termasuk dalam seleksi fitur menggunakan pendekatan filter adalah *Fisher Criterion*. Oleh karena itu, dipilih metode *Fisher Criterion* dalam menyeleksi fitur terbaik. Metode *Fisher Criterion* di dapat dengan menggunakan persamaan berikut,

$$F(x_j) = \frac{n_{+1}n_{-1}}{n_{+1}n_{-1} + n_{j,+1}\sigma_{j,+1}^2 + n_{j,-1}\sigma_{j,-1}^2} \frac{(\mu_{j,+1} - \mu_{j,-1})^2}{n_{j,+1}\sigma_{j,+1}^2 + n_{j,-1}\sigma_{j,-1}^2} \quad (4)$$

dimana n_{+1} adalah jumlah data pada fitur tertentu yang termasuk pada kelas epitel, n_{-1} adalah jumlah data pada fitur tertentu yang termasuk pada kelas radang, $\mu_{j,+1}$ adalah rata-rata data pada fitur ke-j yang termasuk pada kelas epitel, $\mu_{j,-1}$ adalah rata-rata data pada fitur ke-j yang termasuk pada kelas radang, $\sigma_{j,+1}^2$ adalah kuadrat dari standar deviasi data pada fitur ke-j yang termasuk pada kelas epitel, $\sigma_{j,-1}^2$ adalah kuadrat dari

standar deviasi data pada fitur ke-j yang termasuk pada kelas radang. Algoritma ini menghasilkan output berupa ranking setiap fitur.

2. Backpropagation

Setelah mendapatkan ranking setiap fitur langkah selanjutnya adalah melakukan proses klasifikasi. Pada penelitian ini proses klasifikasi dilakukan dengan menggunakan *Backpropagation*. Pada penelitian ini Jumlah *hidden layer* yang digunakan berjumlah 9, dengan jumlah epoh adalah sebanyak 1000. Gambar 7. Menunjukkan arsitektur jaringan syaraf pada penelitian ini.

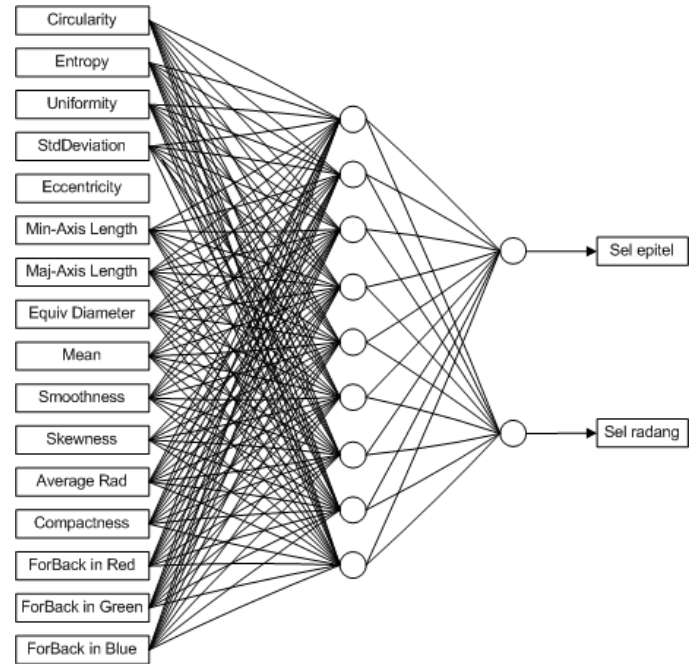


Fig. 7. Arsitektur jaringan saraf pada penelitian ini.

3. K-Fold Cross Validation

K-Fold Cross Validation adalah sebuah metode yang membagi himpunan contoh secara acak menjadi k himpunan bagian (subset). K-Fold Cross melakukan iterasi sebanyak k kali untuk data pelatihan dan pengujian. Pada setiap iterasi, satu subset digunakan untuk pengujian sedangkan subset sisanya digunakan untuk pelatihan.

Data awal dibagi menjadi k subset secara acak dengan ukuran subset yang hampir sama dengan mempertahankan perbandingan antar kelas. Pada iterasi pertama, subset satu menjadi data pengujian sedangkan subset lainnya menjadi data

TABLE IV. RANGKING FITUR MENGGUNAKAN FISHER CRITERION

| Rangking | Fitur |
|----------|---|
| 1 | Minor Axis Length |
| 2 | Average Radius |
| 3 | Equivalent Diameter |
| 4 | Major Axis Length |
| 5 | Uniformity |
| 6 | Foreground-Background contrast in red |
| 7 | Compactness |
| 8 | Circularity |
| 9 | Eccentricity |
| 10 | Skewness (Third Moment) |
| 11 | Foreground-Background contrast in green |
| 12 | Foreground-Background contrast in blue |
| 13 | Stdev (Second Moment) |
| 14 | Smoothness |
| 15 | Entropy |
| 16 | Mean (First Moment) |

pelatihan. Pada iterasi kedua, subset kedua digunakan sebagai data pengujian dan subset lainnya sebagai data pelatihan, dan seterusnya hingga seluruh subset digunakan sebagai data pengujian.

Pada penelitian ini digunakan metode *k-fold cross validation* dengan nilai $k=10$. Data yang digunakan dibagi menjadi sepuluh subset secara acak yang masing-masing subset memiliki jumlah instance dan perbandingan jumlah kelas yang sama. Pembagian subset untuk setiap data tergantung pada jumlah instance dan jumlah kelas masing-masing. Pembagian data ini digunakan pada proses iterasi klasifikasi. Iterasi dilakukan sebanyak sepuluh sesuai dengan nilai k . Pada setiap iterasi, satu subset digunakan untuk pengujian sedangkan subset-subset lainnya digunakan untuk pelatihan.

III. HASIL

Dataset yang digunakan adalah sebanyak 321 data yang diambil dari 20 citra Pap smear yang berbeda. Adapun proses klasifikasi pada data yang digunakan telah melalui proses *assessment* dari dokter Spesialis Patologi Anatomi. Dataset yang digunakan dibagi menjadi 2 klasifikasi yaitu sel radang dan sel epitel. Data sel epitel yang digunakan adalah sebanyak 177 data dan 144 data merupakan sel radang. Hasil dari proses ekstraksi fitur kemudian dilakukan normalisasi data karena adanya perbedaan rentang nilai pada tiap fitur dengan menggunakan normalisasi *min-max* berikut,

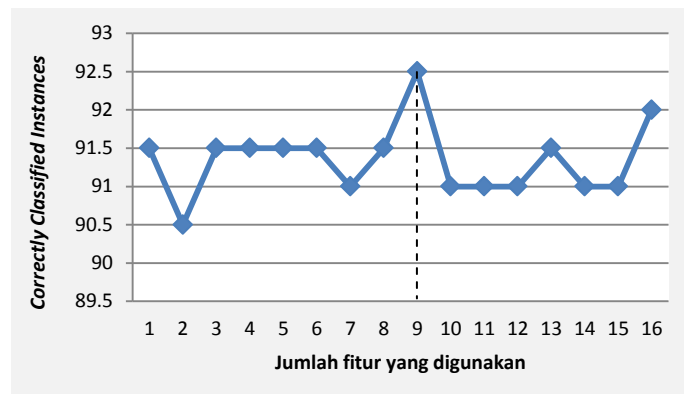
$$N(i) = \frac{F(i) - \min(F)}{\max(F) - \min(F)} \in D_F \quad (3)$$

dengan F adalah index fitur yang di gunakan, dan i adalah index data pada dimensi (D_F).

Output dari algoritma *Fisher Criterion* adalah berupa rangking fitur dimana semakin tinggi rangking sebuah fitur, maka tingkat signifikannya terhadap hasil klasifikasi semakin tinggi. Berikut Tabel 4 menunjukkan hasil rangking terhadap fitur yang digunakan, sedangkan Gambar 8 memperlihatkan hasil *recognition rates* yang dihasilkan oleh metode *Backpropagation*.

TABLE V. RANGKING FITUR DENGAN FSS

| Rangking | Fitur |
|----------|---|
| 1 | Minor Axis Length |
| 2 | Equivalent Diameter |
| 3 | Average Radius |
| 4 | Uniformity |
| 5 | Major Axis Length |
| 6 | Circularity |
| 7 | Compactness |
| 8 | Foreground-Background contrast in red |
| 9 | Eccentricity |
| 10 | Entropy |
| 11 | Stdev (Second Moment) |
| 12 | Smoothness |
| 13 | Skewness (Third Moment) |
| 14 | Foreground-Background contrast in blue |
| 15 | Mean (First Moment) |
| 16 | Foreground-Background contrast in green |

Fig. 8. *Recognition rates* dari metode *Backpropagation* terhadap *Fisher Criterion*

IV. DISKUSI

Dari Gambar 8 dapat disimpulkan bahwa secara umum nilai akurasi data mengalami kenaikan jika dilakukan seleksi fitur sebelumnya. Nilai akurasi data jika menggunakan semua fitur adalah 92%, sedangkan jika sebelumnya dilakukan proses seleksi fitur nilai akurasinya meningkat menjadi 92.5 %. Hal itu menunjukkan bahwa tidak semua fitur relevan terhadap hasil klasifikasi. Kesalahan klasifikasi bisa jadi disebabkan karena banyaknya fitur yang kurang relevan terhadap hasil sehingga menurunkan tingkat akurasi data. Proses penghitungan *Recognition rates* menggunakan metode *Backpropagation* dilakukan dengan bantuan *software Weka*.

Sebagai perbandingan hasil yang diperoleh, proses seleksi fitur akan dibandingkan dengan metode lain yakni *Feature Subset Selection* (FSS). Tabel 5 menunjukkan hasil perbandingan dengan menggunakan FSS. Gambar 9 menunjukkan nilai akurasi tertinggi adalah ketika menggunakan 9 fitur yaitu 92.5%. Hal ini berarti fitur yang digunakan adalah fitur dengan rangking 1 sampai 9. Hal ini menunjukkan kesamaan fitur yang terpilih antara menggunakan *Fisher Criterion* maupun dengan *Feature Subset Selection* meskipun urutan rangkingnya berbeda. Berikut Tabel 6 menunjukkan fitur yang terpilih menggunakan seleksi fitur

baik menggunakan *Fisher Criterion* maupun dengan *Feature Subset Selection*.

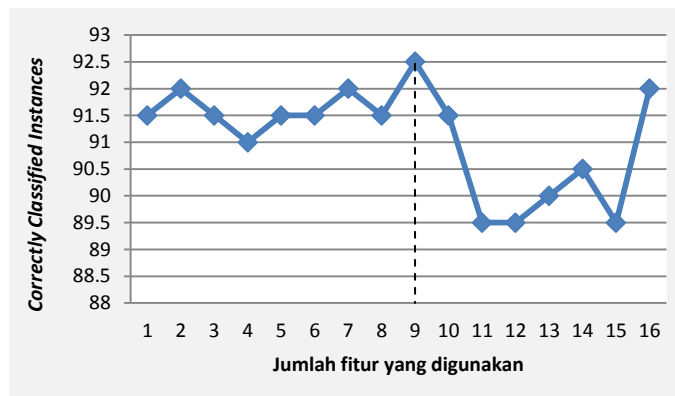


Fig. 9. Recognition rates dari metode *Backpropagation* terhadap *Feature Subset Selection*

Tabel 6 menunjukkan bahwa walaupun terdapat perbedaan urutan ranking fitur antara *Fisher Criterion* dengan *Feature Subset Selection*, namun kedua metode tersebut memiliki kesamaan fitur apa saja yang terbukti signifikan untuk mengklasifikasikan sel radang dengan sel epitel. Dari Tabel 6 dapat disimpulkan bahwa untuk melakukan klasifikasi sel epitel dengan sel radang dapat dilakukan dengan 9 fitur yang dapat dikelompokkan menjadi berikut:

1. Fitur bentuk: *Minor Axis Length*, *Average Radius*, *Equivalent Diameter*, *Major Axis Length*, *Compactness*, *Circularity* dan *Eccentricity*.
2. Fitur Tekstur: *Uniformity*.
3. Fitur Intensitas warna: *Foreground-Background contrast in red*.

TABLE VI. FITUR YANG TERPILIH

| Rangking | Fisher Criterion | Feature Subset Selection |
|----------|--|--|
| 1 | <i>Minor Axis Length</i> | <i>Minor Axis Length</i> |
| 2 | <i>Average Radius</i> | <i>Equivalent Diameter</i> |
| 3 | <i>Equivalent Diameter</i> | <i>Average Radius</i> |
| 4 | <i>Major Axis Length</i> | <i>Uniformity</i> |
| 5 | <i>Uniformity</i> | <i>Major Axis Length</i> |
| 6 | <i>Foreground-Background contrast in red</i> | <i>Circularity</i> |
| 7 | <i>Compactness</i> | <i>Compactness</i> |
| 8 | <i>Circularity</i> | <i>Foreground-Background contrast in red</i> |
| 9 | <i>Eccentricity</i> | <i>Eccentricity</i> |

Pada penelitian ini, 9 fitur yang diusulkan sudah terbukti mampu mengklasifikasikan sel radang dengan sel epitel. Keakuratan nilai klasifikasi yang didapat menggunakan *backpropagation* mencapai 92,5%. Pada penelitian yang lain [12] dijelaskan bahwa kriteria *morphometric* seharusnya mencakup 43 fitur untuk klasifikasi antara sel normal maupun sel kanker. Penggunaan 43 fitur tersebut akan diinvestigasi apakah mampu meningkatkan nilai akurasi klasifikasi sel radang dengan sel epitel pada penelitian selanjutnya.

V. KESIMPULAN

Selain mampu mengurangi waktu komputasi, seleksi fitur terbukti mampu meningkatkan tingkat akurasi dalam proses klasifikasi. Keakuratan nilai klasifikasi yang didapat pada penelitian ini dengan seleksi fitur mencapai 92,5%. Penelitian ini juga menyimpulkan bahwa untuk melakukan proses klasifikasi antara sel radang dan sel epitel dapat dilakukan menggunakan 9 fitur yaitu: *Minor Axis Length*, *Average Radius*, *Equivalent Diameter*, *Major Axis Length*, *Uniformity*, *Foreground-Background contrast in red*, *Compactness*, *Circularity*, dan *Eccentricity*. Dari hasil penelitian ini dapat digunakan untuk penelitian selanjutnya untuk meningkatkan keakuratan deteksi nuclei pada citra Pap smear.

REFERENCES

- [1] Plissiti, ME., Nikou, C., and Charchanti, A., "Automated detection of Cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering," *IEEE Trans on Information technology in Biomedicine*, vol. 15, no 2, March. 2011, pp. 233-241.
- [2] Muhimmah, I., Kurniawan, R., Indrayanti, "Automatic Epithelial Cells Detection of Pap smears images using Fuzzy C-Means Clustering", *2012 4th International Conference on Bioinformatics and Biomedical Technology*, pp. 122-127, 2012.
- [3] Lin, C. H., Chan, Y. K., and Chen, C. C., "Detection and segmentation of cervical cell cytoplasm and nucleus," *Int. J. Imaging Syst. Technol.*, vol. 19, no. 3, pp. 260–270, 2009.
- [4] Yang-Mao, S. F., Chan, Y. K., and Chu, Y. P., "Edge enhancement nucleus and cytoplasm contour detector of cervical smear images," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 38, no. 2, pp. 353–366, Apr. 2008.
- [5] Garrido, A., N. Pérez de la Blanca, "Applying deformable templates for cell image segmentation", *Pattern Recognition*, vol. 33, 2000, pp. 821-832
- [6] Bak, E., Najarian, K., and Brockway, J. P., "Efficient segmentation framework of cell images in noise environments," in *Proc. 26th Int. Conf. IEEE Eng. Med. Biol.*, Sep., 2004, vol. 1, pp. 1802–1805
- [7] Muhimmah, I., Kurniawan, R., Indrayanti (2012) Automated Cervical Cell Nuclei Segmentation using Morphological Operation and Watershed Transformation. In *IEEE International Conference on Computational Intelligence and Cybernetics* 2012.
- [8] A.K. Jain, D. Zongker, Feature selection: evaluation, application and small sample performance, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (2) (1997) 153–158.
- [9] M. E. Plissiti, C. Nikou, and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images," *Pattern Recognition Letters*, vol. 32, no. 6, pp. 838–853, 2011.
- [10] Drezek R, Guillaud M, Collier T, Boiko I, Malpica A, Macaulay C, Follen M, Richards-Kortum R, "Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, DNA content, and chromatin texture", *J. Biomed. Opt.* 8, 7 (2003);
- [11] L.A. West, R. Swartz, D. Cox, I.V. Boiko, A. Malpica, C. MacAulay, M. Follen, "Cytometric features of cell nuclei of adenocarcinoma in situ and invasive adenocarcinoma of the cervix," *Am J Obstet Gynecol*, vol.187, pp.1566-1573, 2002.
- [12] Guillaud, M., Cox, D., Adler-Storthz, K., Malpica, A., Staerke, G., Matisic, J., van Niekerk, D., Poulin, N., Follen, M., and MacAulay, C., "Exploratory Analysis of Quantitative Histopathology of Cervical Intraepithelial Neoplasia: Objectivity, Reproducibility, Magnificancy-Associated Changes, and Human Papillomavirus". *Cytometry Part A* vol. 60A, 2004, pp.8